

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/925,122A

DATE: 12/13/2001

TIME: 15:07:32

Input Set : N:\paola\09925122A.txt

Output Set: N:\CRF3\12132001\I925122A.raw

SEQUENCE LISTING

C--> 3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: Bandman, Olga
 6 Guegler, Karl J.
 7 Lal, Preeti

C--> 9 (ii) TITLE OF INVENTION: SH3-CONTAINING PROTEINS

12 (iii) NUMBER OF SEQUENCES: 5

14 (iv) CORRESPONDENCE ADDRESS:

15 (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
 16 (B) STREET: 3174 Porter Dr.
 17 (C) CITY: Palo Alto
 18 (D) STATE: CA
 19 (E) COUNTRY: USA
 20 (F) ZIP: 94304

22 (v) COMPUTER READABLE FORM:

23 (A) MEDIUM TYPE: Diskette
 24 (B) COMPUTER: IBM Compatible
 25 (C) OPERATING SYSTEM: DOS
 26 (D) SOFTWARE: FastSEQ for Windows Version 2.0

28 (vi) CURRENT APPLICATION DATA:

C--> 29 (A) APPLICATION NUMBER: US/09/925,122A

C--> 30 (B) FILING DATE: 08-Aug-2001

32 (vii) PRIOR APPLICATION DATA:

33 (A) APPLICATION NUMBER: 09/294,545
 34 (B) FILING DATE: 1997-11-13

36 (viii) ATTORNEY/AGENT INFORMATION:

37 (A) NAME: Billings, Lucy J.
 38 (B) REGISTRATION NUMBER: 36,749
 39 (C) REFERENCE/DOCKET NUMBER: PF-0419 US

41 (ix) TELECOMMUNICATION INFORMATION:

42 (A) TELEPHONE: 650-855-0555
 43 (B) TELEFAX: 650-845-4166

45 (2) INFORMATION FOR SEQ ID NO: 1:

47 (i) SEQUENCE CHARACTERISTICS:

48 (A) LENGTH: 265 amino acids
 49 (B) TYPE: amino acid
 50 (C) STRANDEDNESS: single
 51 (D) TOPOLOGY: linear

53 (vii) IMMEDIATE SOURCE:

54 (A) LIBRARY: BRAITUT03
 55 (B) CLONE: 865744

57 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

59 Met Lys Arg Thr Val Ser Asp Asn Ser Leu Ser Asn Ser Arg Gly Glu
 60 1 5 10 15
 61 Gly Lys Pro Asp Leu Lys Phe Gly Gly Lys Ser Lys Gly Lys Leu Trp
 62 20 25 30
 63 Pro Phe Ile Lys Lys Asn Lys Gly Ala Thr Pro Glu Asp Phe Ser Asn

ENTERED

TECH CENTER 1600/2900

JAN 18 2002

RECEIVED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/925,122A

DATE: 12/13/2001

TIME: 15:07:32

Input Set : N:\paola\09925122A.txt

Output Set: N:\CRF3\12132001\I925122A.raw

```

64          35          40          45
65 Leu Pro Pro Glu Gln Arg Arg Lys Lys Leu Gln Gln Lys Val Asp Glu
66      50          55          60
67 Leu Asn Lys Glu Ile Gln Lys Glu Met Asp Gln Arg Asp Ala Ile Thr
68 65          70          75          80
69 Lys Met Lys Asp Val Tyr Leu Lys Asn Pro Gln Met Gly Asp Pro Ala
70          85          90          95
71 Ser Leu Asp His Lys Leu Ala Glu Val Ser Gln Asn Ile Glu Lys Leu
72          100          105          110
73 Arg Val Glu Thr Gln Lys Phe Glu Ala Trp Leu Ala Glu Val Glu Gly
74          115          120          125
75 Arg Leu Pro Ala Arg Asn Glu Gln Ala Arg Arg Gln Ser Gly Leu Tyr
76          130          135          140
77 Asp Ser Gln Asn Pro Pro Thr Val Asn Asn Cys Ala Gln Asp Arg Glu
78 145          150          155          160
79 Ser Pro Asp Gly Ser Tyr Thr Glu Glu Gln Ser Gln Glu Ser Glu Met
80          165          170          175
81 Lys Val Leu Ala Thr Asp Phe Asp Asp Glu Phe Asp Asp Glu Glu Pro
82          180          185          190
83 Leu Pro Ala Ile Gly Thr Cys Lys Ala Leu Tyr Thr Phe Glu Gly Gln
84          195          200          205
85 Asn Glu Gly Thr Ile Ser Val Val Glu Gly Glu Thr Leu Tyr Val Ile
86          210          215          220
87 Glu Glu Asp Lys Gly Asp Gly Trp Thr Arg Ile Arg Arg Asn Glu Asp
88 225          230          235          240
89 Glu Glu Gly Tyr Val Pro Thr Ser Tyr Val Glu Val Cys Leu Asp Lys
90          245          250          255
91 Asn Ala Lys Gly Ala Lys Thr Tyr Ile
92          260          265

```

94 (2) INFORMATION FOR SEQ ID NO: 2:

96 (i) SEQUENCE CHARACTERISTICS:

97 (A) LENGTH: 1459 base pairs

98 (B) TYPE: nucleic acid

99 (C) STRANDEDNESS: single

100 (D) TOPOLOGY: linear

102 (vii) IMMEDIATE SOURCE:

103 (A) LIBRARY: BRAITUT03

104 (B) CLONE: 865744

106 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

108 AGTAAAAGCA GCCGAATCAA TTGATCAGAA AAATGATTCA CAGCTGGTAA TAGAAGCTTA      60
109 TAAATCAGGG TTTGAGCCTC CTGGAGACAT TGAATTTGAG GATTACACTC AGCCAATGAA      120
110 GCGCACTGTG TCAGATAACA GCCTTTCAAA TTCCAGAGGA GAAGGCAAAC CAGACCTCAA      180
111 ATTTGGTGGC AAATCCAAAG GAAAGTTATG GCCGTTTCATC AAAAAAATA AGGGTGCAAC      240
112 ACCGGAGGAT TTCAGCAACC TCCCACCTGA ACAAAGAAGG AAAAAGCTGC AGCAGAAAGT      300
113 CGATGAGTTA AATAAAGAAA TTCAGAAGGA GATGGATCAA AGAGATGCCA TAACAAAAT      360
114 GAAAGATGTC TACCTAAAGA ATCCTCAGAT GGGAGACCCA GCCAGTTTGG ATCACAAATT      420
115 AGCAGAAGTC AGCCAAAATA TAGAGAAACT GCGAGTAGAG ACCCAGAAAT TTGAGGCCTG      480
116 GCTGGCTGAG GTTGAAGGCC GGCTCCCAGC ACGCAACGAG CAGGCGCGCC GGCAGAGCGG      540
117 ACTGTACGAC AGCCAGAACC CACCCACAGT CAACAACTGC GCCCAGGACC GTGAGAGCCC      600

```

RAW SEQUENCE LISTING

DATE: 12/13/2001

PATENT APPLICATION: US/09/925,122A

TIME: 15:07:32

Input Set : N:\paola\09925122A.txt

Output Set: N:\CRF3\12132001\I925122A.raw

```

118 AGATGGCAGT TACACAGAGG AGCAGAGTCA GGAGAGTGAG ATGAAGGTGC TGGCCACGGA      660
119 TTTTGACGAC GAGTTTGATG ATGAGGAGCC CCTCCCTGCC ATAGGGACGT GCAAAGCTCT      720
120 CTACACATTT GAAGGTCAGA ATGAAGGAAC GATTTCGTA GTTGAAGGAG AAACATTGTA      780
121 TGTCATAGAG GAAGACAAAG GCGATGGCTG GACCCGCATT CGGAGAAATG AAGATGAAGA      840
122 GGGTTATGTC CCCACTTCAT ATGTCGAAGT CTGTTTGGAC AAAAATGCCA AAGGTGCTAA      900
123 GACTTATATT TAATACCATA AAAAAAAAAA ACTTAAAAAA AATGGAGTTG TTTCTCCCCA      960
124 CAACCGTGAC TGTTACAGGC AGTTCCTCAA GAGACTGGCT GGCAAGCACC ATAATGCACG     1020
125 TTCTCCTGTA GTCTCACGTG GACTTCAGGG TCCGGGCACC TGAATTGCCT TGTCTAGTTT     1080
126 GGGCTGTAAT CAAGTTTCAC TTGCTGATGA AATTTTATGT GGAAAGCTGC CAACCGCCAA     1140
127 CTTACAGCTA TGTCATTCAA AATCTGATAA ACATTTCTTC TTTTGGCGGT ATCTGTAGAT     1200
128 TAAAAAAGAA GTTGCAATTG AGCTTCTCAT CTTTCTGAAT TTAAAAGCCG GCACGCATCA     1260
129 TGCAGGTGCC AAAGACTTCC CTAATCTTGT TTATATCTAG TATCCACCAT ACATGAGCT     1320
130 ACATTAGGTG GTTACAGATT GTAACCTAAT AAAGTGAAGT GTGTTAGTTT GTTAAATTGG     1380
131 ATACTCATTG ACTTGGGGAG GAGTCACAAG TGAAATACCA TCTCTTTCTT GACTAAAGCG     1440
132 GTAAATAAGG TTCTTATTG                                     1459

```

134 (2) INFORMATION FOR SEQ ID NO: 3:

136 (i) SEQUENCE CHARACTERISTICS:

137 (A) LENGTH: 175 amino acids

138 (B) TYPE: amino acid

139 (C) STRANDEDNESS: single

140 (D) TOPOLOGY: linear

142 (vii) IMMEDIATE SOURCE:

143 (A) LIBRARY: PROSNOT20

144 (B) CLONE: 1816529

146 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

148 Met Lys Asp Val Tyr Glu Lys Thr Pro Gln Met Gly Asp Pro Ala Ser
149 1 5 10 15
150 Leu Glu Pro Gln Ile Ala Glu Thr Leu Ser Asn Ile Glu Arg Leu Lys
151 20 25 30
152 Leu Glu Val Gln Lys Tyr Glu Ala Trp Leu Ala Glu Ala Glu Ser Arg
153 35 40 45
154 Val Leu Ser Asn Arg Gly Asp Ser Leu Ser Arg His Ala Arg Pro Pro
155 50 55 60
W--> 156 Xaa Pro Pro Ala Ser Ala Pro Pro Asp Ser Ser Ser Asn Ser Ala Ser
157 65 70 75 80
158 Gln Asp Thr Lys Glu Ser Ser Glu Glu Pro Pro Ser Glu Glu Ser Gln
159 85 90 95
160 Asp Thr Pro Ile Tyr Thr Glu Phe Asp Glu Asp Phe Glu Glu Glu Pro
161 100 105 110
162 Thr Ser Pro Ile Gly His Cys Val Ala Ile Tyr His Phe Glu Gly Ser
163 115 120 125
164 Ser Glu Gly Thr Ile Ser Met Ala Glu Gly Glu Asp Leu Ser Leu Met
165 130 135 140
166 Glu Glu Asp Lys Gly Asp Gly Trp Thr Arg Val Arg Arg Lys Glu Gly
167 145 150 155 160
168 Gly Glu Gly Tyr Val Pro Thr Ser Tyr Leu Arg Val Thr Leu Asn
169 165 170 175

```

171 (2) INFORMATION FOR SEQ ID NO: 4:

173 (i) SEQUENCE CHARACTERISTICS:

RAW SEQUENCE LISTING

DATE: 12/13/2001

PATENT APPLICATION: US/09/925,122A

TIME: 15:07:32

Input Set : N:\paola\09925122A.txt

Output Set: N:\CRF3\12132001\I925122A.raw

```

174      (A) LENGTH: 773 base pairs
175      (B) TYPE: nucleic acid
176      (C) STRANDEDNESS: single
177      (D) TOPOLOGY: linear
179      (vii) IMMEDIATE SOURCE:
180          (A) LIBRARY: PROSNOT20
181          (B) CLONE: 1816529
183      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
185      ATGAACCGTG CACCCTNCGA CAGCAGTCTG GGCACCCCTT ACGGATGGAC GGNCTGAACT      60
186      CCGAGGNCCG GGTGCGAGCC GCACCAAGCG CTGGNCTTTT GGCAAGAAGA ACAAGACAGT      120
187      GGTGACCGAG GATTTTAGCC ACTTGCCCCC AGAGCAGCAG CGAAAACGGC TTCAAACAGCA      180
188      GTTGAAGAA CGCAGTCGTG AACTTCAGAA GGAGGTTGAC CAGAGGGAAG CCCTAAAGAA      240
189      AATGAAGGAT GTCTATGAGA AGACACTCA GATGGGGGAC CCCGCCAGCT TGGAGCCCCA      300
190      GATCGCTGAA ACCCTGAGCA ACATTGAACG GCTGAAATTG GAAGTGCAGA AGTATGAGGC      360
191      GTGGCTGGCA GAAGCTGAAA GTCGAGTCTT TAGCAACCGG GGAGACAGCC TGAGCCGGCA      420
192      CGCCCGGCCT CCCGANCCCC CCGCTAGCGC CCCGCCAGAC AGCAGCAGCA ACAGCGCATC      480
193      ACAGGACACC AAGGAGAGCT CTGAAGAGCC TCCCTCAGAA GAGAGCCAGG ACACCCCAT      540
194      TTACACGGAG TTTGATGAGG ATTTGAGGGA GGAACCCACA TCCCCCATAG GTCACTGTGT      600
195      GGCCATCTAC CACTTTGAAG GGTCCAGCGA GGGCACTATC TCTATGGCCG AGGGTGAAGA      660
196      CCTCAGTCTT ATGGAAGAAG ACAAAGGGGA CGGCTGGACC CGGGTCAGGC GGAAAGAGGG      720
197      AGGCGAGGGC TACGTGCCCA CCTCTACCT CCGAGTCACG CTCAATTGAA CCC      773
199      (2) INFORMATION FOR SEQ ID NO: 5:
201          (i) SEQUENCE CHARACTERISTICS:
202              (A) LENGTH: 237 amino acids
203              (B) TYPE: amino acid
204              (C) STRANDEDNESS: single
205              (D) TOPOLOGY: linear
207          (vii) IMMEDIATE SOURCE:
208              (A) LIBRARY: GenBank
209              (B) CLONE: 1255033
211          (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
213      Lys Ile His Cys Phe Arg Ser Leu Lys Arg Gly Gly Val Thr Pro Glu
214      1      5      10      15
215      Asp Phe Ser Asn Phe Pro Pro Glu Gln Arg Arg Lys Lys Leu Gln Gln
216      20      25      30
217      Lys Val Asp Asp Leu Asn Arg Glu Ile Gln Lys Glu Thr Asp Gln Arg
218      35      40      45
219      Asp Ala Ile Thr Lys Met Lys Asp Val Tyr Leu Lys Asn Pro Gln Met
220      50      55      60
221      Gly Asp Pro Ala Ser Leu Asp Gln Lys Leu Thr Glu Val Thr Gln Asn
222      65      70      75      80
223      Ile Glu Lys Leu Arg Leu Glu Ala Gln Lys Phe Glu Ala Trp Leu Ala
224      85      90      95
225      Glu Val Glu Gly Arg Leu Pro Ala Arg Ser Glu Gln Ala Arg Arg Gln
226      100     105     110
227      Ser Gly Leu Tyr Asp Gly Gln Thr His Gln Thr Val Thr Asn Cys Ala
228      115     120     125
229      Gln Asp Arg Glu Ser Pro Asp Gly Ser Tyr Thr Glu Glu Gln Ser Gln
230      130     135     140

```

RAW SEQUENCE LISTING

DATE: 12/13/2001

PATENT APPLICATION: US/09/925,122A

TIME: 15:07:32

Input Set : N:\paola\09925122A.txt

Output Set: N:\CRF3\12132001\I925122A.raw

231	Glu	Ser	Glu	His	Lys	Val	Leu	Ala	Pro	Asp	Phe	Asp	Asp	Glu	Phe	Asp
232	145					150					155					160
233	Asp	Glu	Glu	Pro	Leu	Pro	Ala	Ile	Gly	Thr	Cys	Lys	Ala	Leu	Tyr	Thr
234					165					170						175
235	Phe	Glu	Gly	Gln	Asn	Glu	Gly	Thr	Ile	Ser	Val	Val	Glu	Gly	Glu	Thr
236				180				185							190	
237	Leu	Ser	Val	Ile	Glu	Glu	Asp	Lys	Gly	Asp	Gly	Trp	Thr	Arg	Ile	Arg
238			195					200					205			
239	Arg	Asn	Glu	Asp	Glu	Glu	Gly	Tyr	Phe	Pro	Thr	Ser	Tyr	Val	Glu	Val
240		210					215					220				
241	Tyr	Leu	Asp	Lys	Asn	Ala	Lys	Gly	Ala	Lys	Thr	Tyr	Ile			
242	225					230					235					

VERIFICATION SUMMARY

DATE: 12/13/2001

PATENT APPLICATION: US/09/925,122A

TIME: 15:07:33

Input Set : N:\paola\09925122A.txt

Output Set: N:\CRF3\12132001\I925122A.raw

L:3 M:220 C: Keyword misspelled or invalid format, [(1) GENERAL INFORMATION:]
L:9 M:220 C: Keyword misspelled or invalid format, [(ii) TITLE OF INVENTION:]
L:29 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:30 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:156 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3